SUGGESTING RELEVANT QUESTIONS FOR A QUERY USING NATURAL LANGUAGE PROCESSING TECHNIQUES

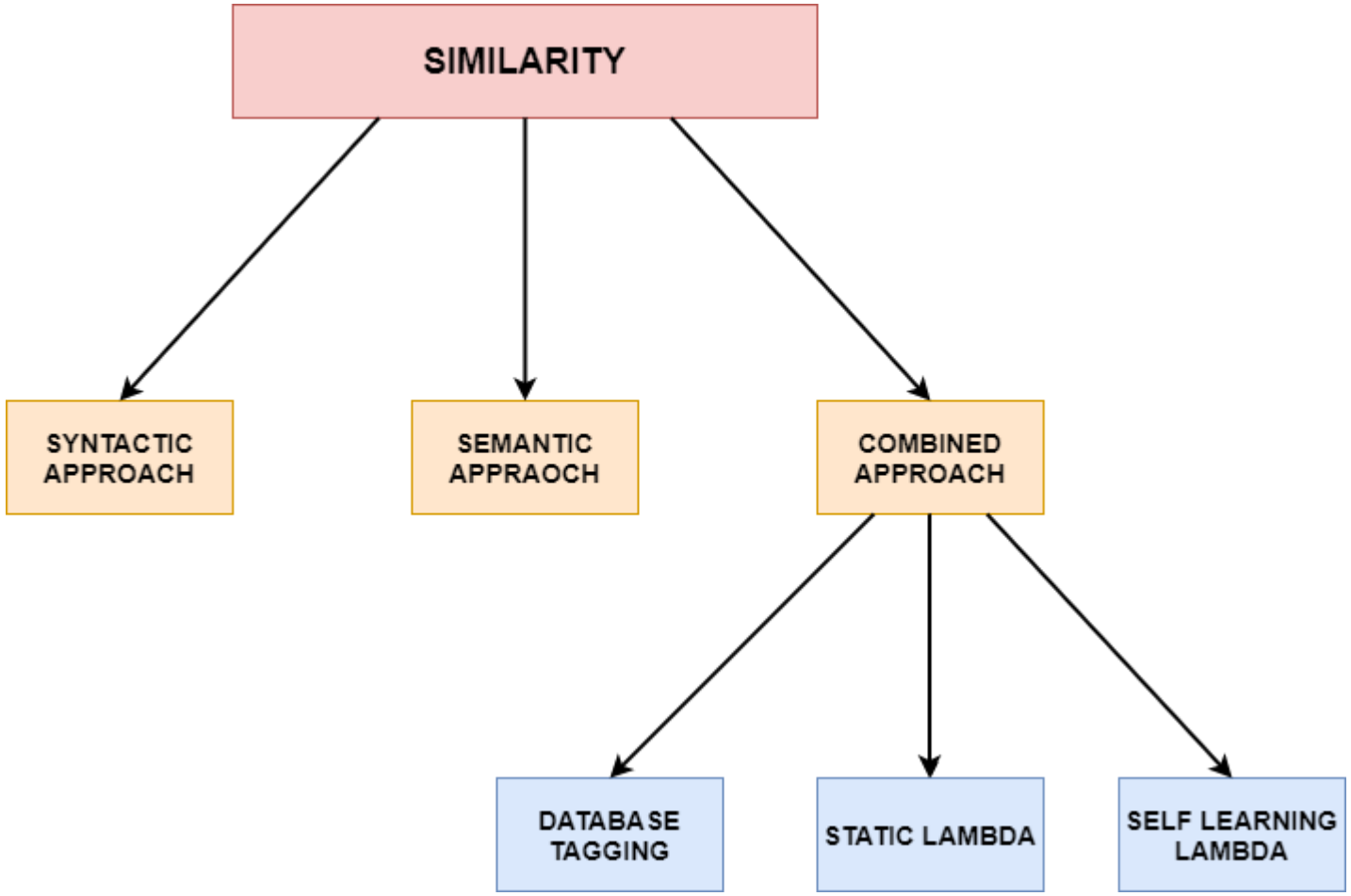
Internal Guide : Dr. A.S. Ghotkar
(*Associate Professor, Dept of CE, PICT*)
External Guide : Mr. Nikhil Malhotra
(*Global Head, Maker's Lab, Tech Mahindra*)

Hrushabh Hirudkar	71700888G
Anuj Kanetkar	71700929H
Shriniwas Nayak	71701044K

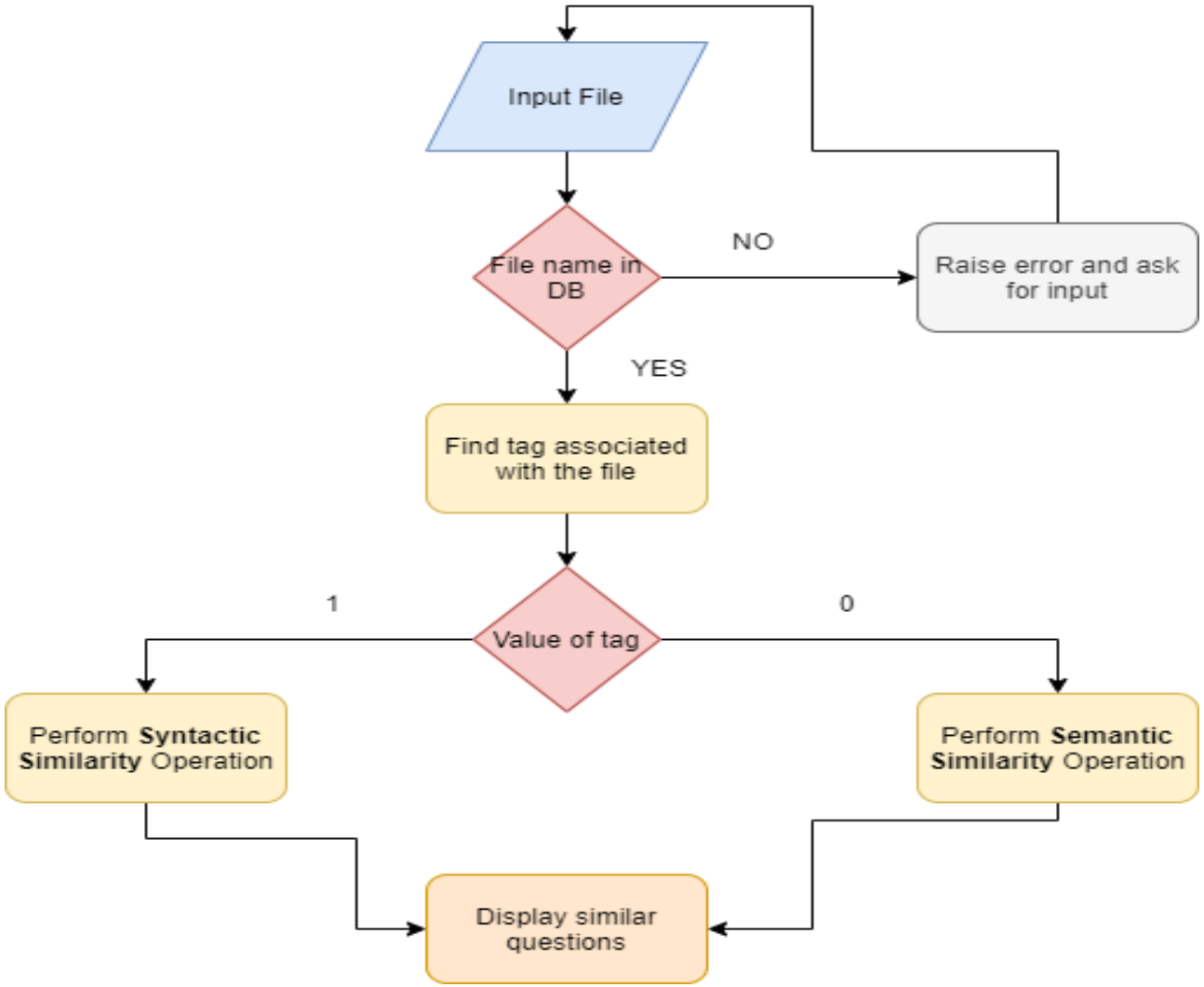
PROBLEM STATEMENT

- We aim to solve the problem by suggesting similar questions to the query entered by the user. The generation of questions involves finding relevant questions from the existing database as well as generating new questions which are not present in the database.

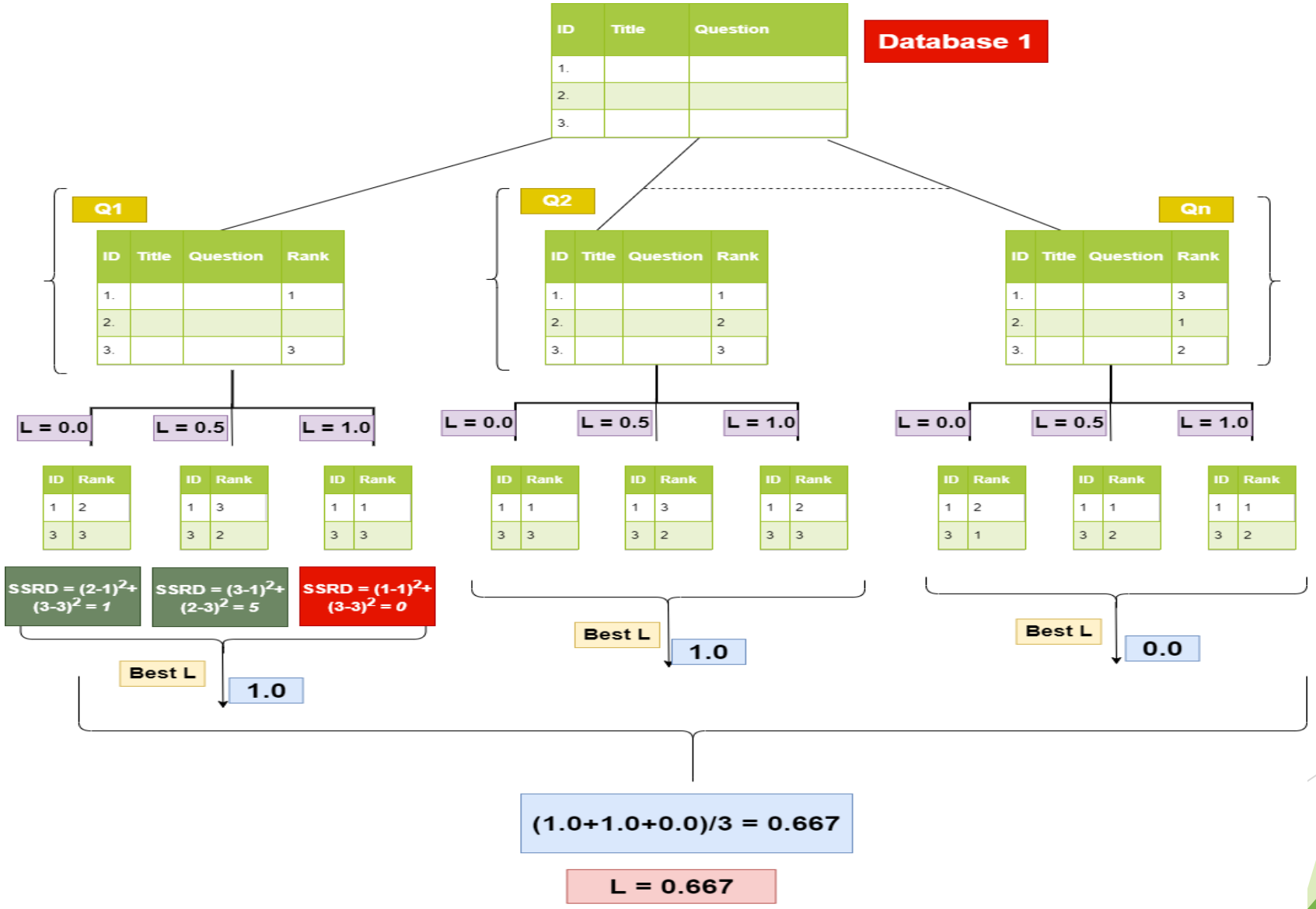
PROJECT HIERARCHY



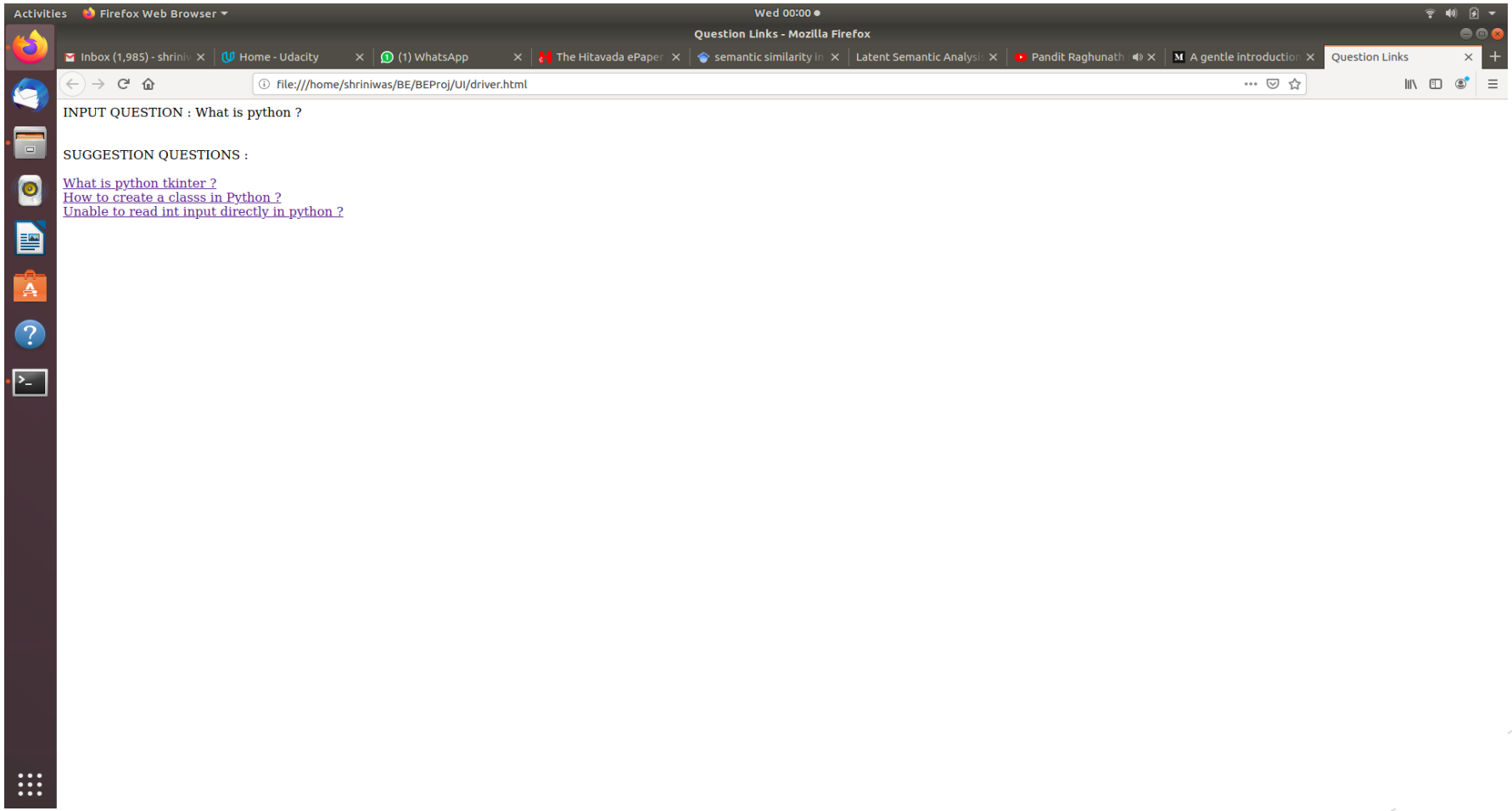
PRODUCT ARCHITECTURE

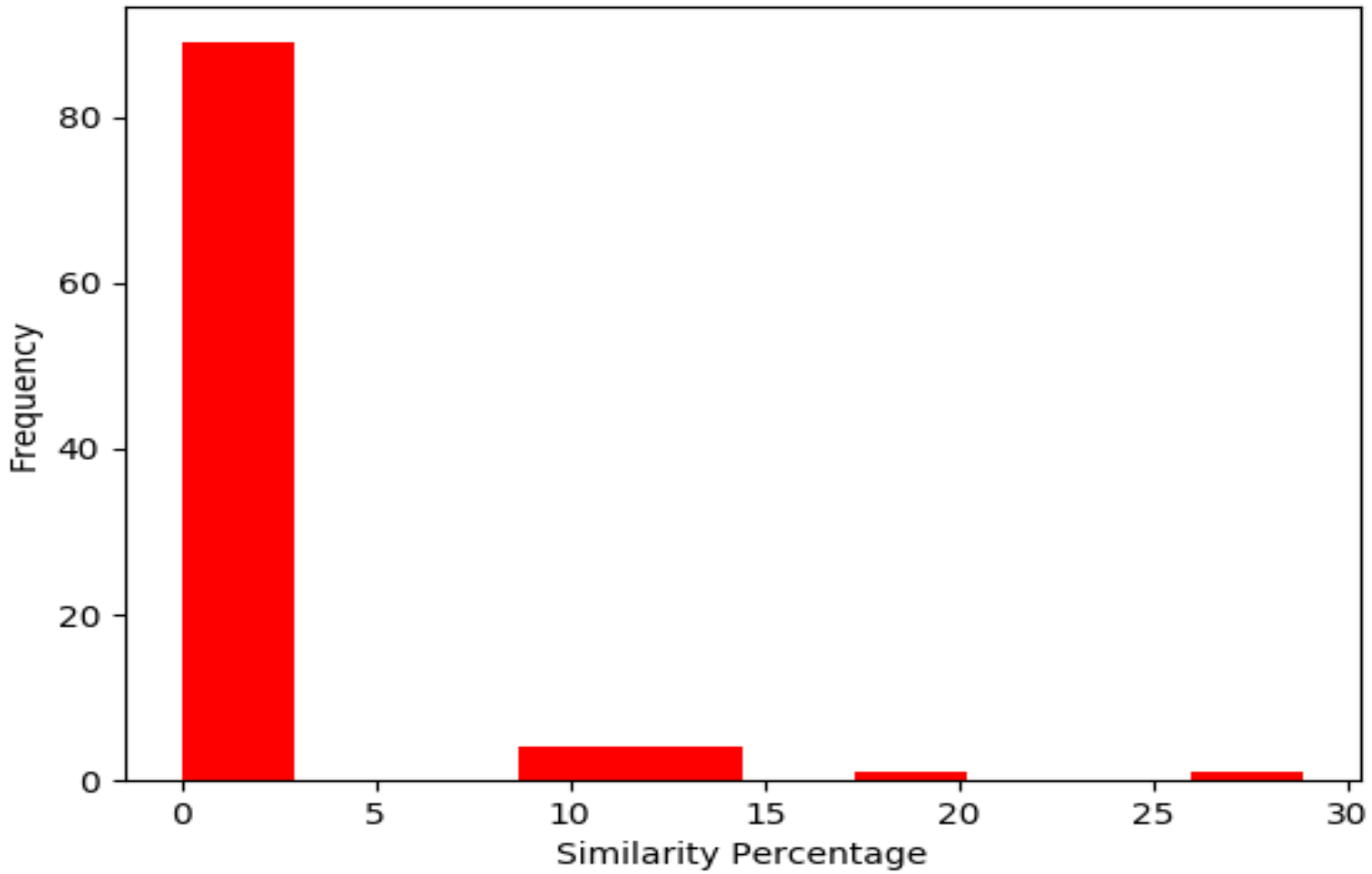


SELF LEARNING LAMBDA ARCHITECTURE



IMPLEMENTATION RESULTS




Frequency
80
60
40
20
0
0
5
10
15
20
25
30
Similarity Percentage

CONCLUSION AND FUTURE SCOPE

- Syntactic similarity has good rejection capacity but fails to accept semantically similar sentences.
- Semantic similarity is good at accepting similar sentences but bad at rejecting the dissimilar ones.
- The future plan for the project can be summarized as follows:
  - Complete the UI implementation (20% work left)
  - Complete the Project report (30% work left)

REFERENCES

[1] Y. Song and D. Roth, “Unsupervised sparse vector densification for shorttext similarity,” inProceedings of the 2015 Conference of the North AmericanChapter of the Association for Computational Linguistics: Human LanguageTechnologies, pp. 1275–1280, 2015.

[2] G. Sidorov, H. G´omez-Adorno, I. Markov, D. Pinto, and N. Loya, “Computingtext similarity using tree edit distance,” in2015 Annual Conference of theNorth American Fuzzy Information Processing Society (NAFIPS) held jointlywith 2015 5th World Conference on Soft Computing (WConSC), pp. 1–4,IEEE, 2015.

[3] C. Schwarz, “lsemantica: A command for text similarity based on latentsemantic analysis,”The Stata Journal, vol. 19, no. 1, pp. 129–142, 2019.

[4] R. Mihalcea, C. Corley, C. Strapparava,et al., “Corpus-based and knowledge-based measures of text semantic similarity,” inAaai, vol. 6, pp. 775–780, 2006.

[5] A. Kashyap, L. Han, R. Yus, J. Sleeman, T. Satyapanich, S. Gandhi, andT. Finin, “Robust semantic text similarity using lsa, machine learning, andlinguistic resources,”Language Resources and Evaluation, vol. 50, no. 1,pp. 125–161, 2016.

[6] A. Islam and D. Inkpen, “Semantic text similarity using corpus-based wordsimilarity and string similarity,”ACM Transactions on Knowledge Discoveryfrom Data (TKDD), vol. 2, no. 2, p. 10, 2008.

[7] W. H. Gomaa and A. A. Fahmy, “A survey of text similarity approaches,”International Journal of Computer Applications, vol. 68, no. 13, pp. 13–18,2013.

[8] X. Wan, “A novel document similarity measure based on earth mover’s dis-tance,”Information Sciences, vol. 177, no. 18, pp. 3718–3730, 2007.

[9] J. Mueller and A. Thyagarajan, “Siamese recurrent architectures for learningsentence similarity,” inThirtieth AAAI Conference on Artificial Intelligence,2016.

[10] A. M. Gliozzo, “Latent semantic analysis for application in a question answersystem,” Apr. 28 2015. US Patent 9,020,810.

THANK YOU!!